Object-Based Multiple Foreground Video Co-Segmentation via Multi-State Selection Graph

Huazhu Fu, Dong Xu, *Senior Member, IEEE*, Bao Zhang, Stephen Lin, *Member, IEEE*, and Rabab Kreidieh Ward, *Fellow, IEEE*

***Abstract*— We present a technique for multiple foreground video co-segmentation in a set of videos. This technique is based on category-independent object proposals. To identify the foreground objects in each frame, we examine the properties of the various regions that reflect the characteristics of foregrounds, considering the intra-video coherence of the foreground as well as the foreground consistency among the different videos in the set. Multiple foregrounds are handled via a multi-state selection graph in which a node representing a video frame can take multiple labels that correspond to different objects. In addition, our method incorporates an indicator matrix that for the first time allows accurate handling of cases with common foreground objects missing in some videos, thus preventing irrelevant regions from being misclassified as foreground objects. An iterative procedure is proposed to optimize our new objective function. As demonstrated through comprehensive experiments, this object-based multiple foreground video co-segmentation method compares well with related techniques that co-segment multiple foregrounds.**

***Index Terms*— Video co-segmentation, multiple foregrounds, object-based segmentation.**

I. Introduction

VIDEO foreground co-segmentation aims at jointly extracting the main common objects present in a given set of videos. Unlike foreground segmentation in the single video case [1], [2], the existence of common foreground objects in multiple videos provides some indication as to which image regions belong to the foreground. However, even with this additional information, much ambiguity still exists in the co-segmentation of general types of videos. Such videos usually contain multiple complex foregrounds that have low contrast with the background. Take for example the two video clips in Fig. 1(a). Each clip contains the same foreground objects. It is shown in Fig. 1(b) that co-segmentation methods based on low-level appearance features may not accurately discriminate between the foreground and background. Also, object-based methods developed for segmentation of a single video do not exploit the joint information between the videos. As a result, the extracted objects may be different from each other as shown in Fig. 1(c).

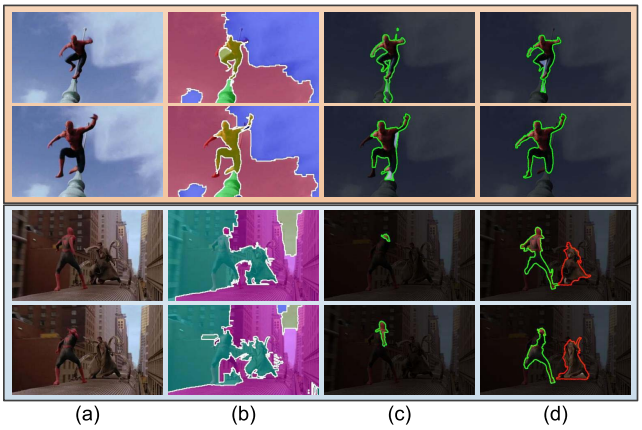


Fig. 1. Video co-segmentation of multiple foreground objects. (a) Two related video clips. (b) Co-segmentation results based on low-level appearance features [3]. (c) Results from object-based video segmentation [2] that does not consider the two videos jointly. (d) Results of our object-based video co-segmentation method.

To address these issues, we develop a video co-segmentation technique that is capable of handling multiple foreground objects in multiple videos. This technique is based on object proposals as the basic element of processing. Object proposals represent regions that are likely to enclose an object according to the structured learning method of [4]. The employment of such mid-level representation of regions has resulted in more robust and meaningful separation of foreground and background regions in images and individual videos [1], [2], [5]–[7]. Our basic video co-segmentation framework is formulated as a graph where nodes represent video frames and take a state that indicates which object proposal has been selected as the foreground. Edges are used between adjacent frames in a video so as to enforce the spatio-temporal smoothness of the trajectory of a foreground

Manuscript received August 30, 2014; revised February 7, 2015 and April 13, 2015; accepted May 21, 2015. Date of publication June 9, 2015; date of current version July 7, 2015. This work was supported by the Singapore A*STAR SERC under Grant 112-148-0003. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Oleg V. Michailovich.

H. Fu is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: hzfu@ntu.edu.sg).

D. Xu is with the School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW 2006, Australia (e-mail: dongxudongxu@gmail.com).

B. Zhang is with the School of Computer Science and Technology, Tianjin University, Tianjin 300072, China (e-mail: zhangbao@tju.edu.cn).

S. Lin is with Microsoft Research, Beijing 100080, China (e-mail: stevelin@microsoft.com).

R. K. Ward is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: rababw@ece.ubc.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2015.2442915

object, while edges between frames of different videos are added to measure foreground consistency.

To handle multiple foregrounds, we introduce the multi-state selection graph (MSG) that enables the selection of multiple states in each node. In the MSG, the basic subgraph is replicated multiple times, with each replicated subgraph representing a particular foreground object. Additional edges between the corresponding nodes of different subgraphs are included to prevent other replicated subgraphs from representing the same object. With this MSG approach, any standard graph model can be converted for multiple state selection. Moreover, the proposed MSG can be optimized by using existing energy minimization techniques.

We further enhance the MSG framework by relaxing the common assumption of object-based segmentation methods that the common object must be present in each image/frame [5], [8] or video [9]. Such a condition is difficult to satisfy in practice, especially for cases with multiple foreground objects. When an object exists, but not in all videos, these object-based methods will segment unrelated regions in place of this missing object. To address this problem, we introduce an indicator matrix in our MSG to correctly deal with missing common objects. Our method yields co-segmentation results that surpass related techniques as shown in Fig. 1(d). For evaluating our multiple foreground video co-segmentation technique, we have built a new dataset with ground truth, which is available online.[1]

II. Related Work

A. Video Co-Segmentation

The few existing methods for video co-segmentation are all based on low-level features. Chen *et al.* [10] detected regions with coherent motion in the videos and then identified a common foreground based on similar chroma and texture feature distributions. Rubio *et al.* [11] separated the foreground and background through an iterative process based on feature matching among video frame regions and spatio-temporal tubes. The low-level appearance models in these methods, however, often do not effectively distinguish between complex foregrounds and backgrounds. Guo *et al.* [12] presented a trajectory co-saliency approach to match actions in a video pair. Its focus is only on common action extraction rather than on foreground object segmentation. The Bag-of-Words (BoW) representation was used for multi-class video co-segmentation by Chiu and Fritz [3]. While BoW features provide more discriminative ability than basic color and texture features, they may be susceptible to appearance variations of a foreground object in different videos, due to factors such as pose change. This problem is illustrated in Fig. 1(b), which shows the co-segmentation results of [3]. It can be seen that the pixel-level features do not provide a representation that is sufficient for relating corresponding regions between the input videos. Guo *et al.* [13] extracted superpixels using a motion-based segmentation method, and introduced the constraint that different parts of the foreground should exist together in all videos, which is evaluated via an iterative constrained clustering. By contrast, the object-based representation [7], [9], [14], [15] can provide greater discriminative ability and robustness, and produces meaningful object-level segments.

[1] https://sites.google.com/site/huazhufu/home/VidCoSeg

B. Object-Based Foreground Segmentation

Object-based techniques are different from the methods based on low-level descriptors, in that they utilize a mid-level region representation designed to capture an object in its entirety. A set of candidate object segments is first generated from an image using object proposal methods [4], [16], [17]. Then, instead of considering segmentation as in the traditional pixel labeling problem, object-based methods perform segmentation by selecting candidates from the set of object proposals [18]. This approach was introduced for co-segmentation of images by Vicente *et al.* [5]. Later, Meng *et al.* [8] used the shortest path algorithm to select a common foreground from object proposals in multiple images. Lee *et al.* [6] took the object proposal regions as foreground candidates for single video segmentation, and also used the objectness score for ranking foreground hypotheses. More recently, the methods for single video segmentation have extended this object-based approach and incorporated the constraint that the foreground appears in every frame. This constraint was formulated within a weighted graph model, with the solution optimized via maximum weight cliques [1], the shortest path algorithm [15], or dynamic programming [2]. Since these segmentation methods deal with a single video rather than the multiple video co-segmentation problem, they do not take advantage of the information from other videos. They also do not address the problem of handling multiple foreground objects. In our work, we utilize a more general co-segmentation graph to formulate the correspondences between different videos and to handle the multiple foreground objects using the MSG model.

C. Multiple Foreground Co-Segmentation

There exist a few co-segmentation methods that can extract multiple foreground objects. Kim *et al.* [19] presented a method based on anisotropic diffusion to identify multiple object classes in a set of images. They later developed a different approach for multiple foreground co-segmentation in images which alternates between foreground modeling and region assignment [20]. In [21], Joulin et al. proposed an energy-based image co-segmentation method that combines spectral and discriminative clustering terms. Mukherjee *et al.* [22] segmented multiple objects from image collections through an analysis of their shared subspace structure. The video co-segmentation method of Chiu and Fritz [3] can also extract multiple foreground objects by learning a global appearance model that connects segments of the same class. However, all these methods cluster the foregrounds into classes based on low-level feature representations. By contrast, object-based techniques operate on the mid-level representation of object proposals but lack an effective way to deal with multiple foregrounds. In our work, we extend the object-based co-segmentation approach to accommodate multiple foregrounds via the proposed

MSG model, where multiple foreground objects in different videos are segmented jointly.

D. Multi-Label Learning

Multi-label learning methods [23], [24] have been used for image/video annotation to propagate image-level labels to different regions. While this may appear similar to our method, there are important differences. In the existing multi-label learning methods, there is a training dataset in which each training image is associated with labels. Various methods based on different models including max-margin [25], nearest neighbors [26], sparse coding [27], and Latent Dirichlet Allocation [28] were used to assign the image-level labels to different regions. In some approaches, the models are applied on the over-segmented regions to predict their labels, and the over-segmented regions with the same assigned label are merged via a post-processing step [29].

In our work, we mainly focus on unsupervised video co-segmentation of foreground objects that are common in a set of videos. This is under the assumption that we only know the number of the foreground objects that are common among the videos and that we do not have a separate training dataset with each training image associated with class labels. While we additionally consider the special case where the objects that exist in each video are known, we concentrate on the problem when these objects are not known (but belong to a known set of candidates with substantial overlap). We aim to find the best candidates for each video from this set of overlapping candidates. This differs from the multi-label learning setting where the instances are often assumed to be independent and non-overlapping. As a result, the existing multi-label learning methods are not suitable for our video co-segmentation task. We therefore propose new Conditional Random Field (CRF) based approaches for video co-segmentation. We are not aware of any video co-segmentation methods that use multi-label learning.

III. Multi-State Selection Graph

In this paper, we introduce a multi-state selection graph (MSG) for co-segmenting multiple common foreground objects in a set of videos. We regard an object in a video to be a foreground if it satisfies the following conditions: 1) The foreground object has a high appearance contrast relative to the background. 2) The trajectory of the foreground object across consecutive video frames is smooth, and the appearance and shape are also similar across the frames. 3) In a video, the foreground object appears in each frame. These three observations are often used in single video segmentation [1], [2], [15]. Moreover, for video co-segmentation, we also introduce an additional constraint on common foreground objects, that they maintain a consistent appearance across different videos. Note that our foreground definition is different from that of traditional moving region separation tasks [30], [31], which deal with surveillance videos captured by static cameras. By contrast, our videos have cluttered backgrounds as well as significant camera motions, which are more challenging than surveillance videos.

We denote a basic subgraph as $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, with a set of nodes $\mathcal{V}$ and edges $\mathcal{E}$. Each node $n \in \mathcal{V}$ can take a state u_n in a discrete space, and has an associated unary energy function $\Psi(u_n)$. An edge $(n, m) \in \mathcal{E}$ connecting nodes n and m has an associated pairwise energy function $\Phi(u_n, u_m)$. The overall energy function of this basic subgraph consists of these two components [32]:

$$E_g(\mathbf{u}) = \sum_{n \in \mathcal{V}} \Psi(u_n) + \sum_{(n,m) \in \mathcal{E}} \Phi(u_n, u_m). \tag{1}$$

The configuration of states $\mathbf{u} = \{u_1, \cdots, u_{|\mathcal{V}|}\}$ can be obtained by minimizing Eq. (1). All existing graph-based algorithms [33]–[36] are based on a common graph definition in which *each node can take only a single state*. In this paper, we extend this definition to consider selecting multiple states for each node under global constraints. Suppose there are K different candidate series $\{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(K)}\}$ to be selected in the graph. For each candidate series $\mathbf{u}^{(k)}$, its solution can be obtained by minimizing the energy of its corresponding basic subgraph $E_g(\mathbf{u}^{(k)})$ in Eq. (1). At the same time, there needs to be minimal correlation among different candidate series. In other words, we seek multiple solutions $\{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(K)}\}$ that are independent as much as possible. To model this independence, we introduce a *diversity term* $\Delta(\mathbf{u}^{(k)}, \mathbf{u}^{(j)})$ between different candidate series, and define the optimization problem for our MSG as follows:

$$\min_{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(K)}} \sum_{k=1}^{K} E_g(\mathbf{u}^{(k)}) + \sum_{k,j=1}^{K} \Delta(\mathbf{u}^{(k)}, \mathbf{u}^{(j)}), \tag{2}$$

where $\{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(K)}\}$ denotes K different candidate series, and $\Delta(\cdot, \cdot)$ represents the diversity term between a pair of candidate series. The MSG is illustrated in Fig. 2(a). To solve the optimization problem for the MSG, we replicate the basic subgraph $\mathcal{G}$ by $K - 1$ times to produce K basic subgraphs in total, one for each candidate series. The diversity term $\Delta(\cdot, \cdot)$ is incorporated as the edges between the corresponding nodes in the basic subgraphs. By linking the basic subgraphs in this manner, they are combined into a unified graph, such that the basic subgraph $\mathcal{G}$ is extended into the MSG $\mathcal{G}'$. We define the energy function for the MSG as follows:

$$\begin{aligned} E_{msg} &= \sum_{k=1}^{K} \left[\sum_{n \in \mathcal{V}} \Psi(u_n^{(k)}) + \sum_{(n,m) \in \mathcal{E}} \Phi(u_n^{(k)}, u_m^{(k)}) \right] \\ &\quad + \sum_{(k,j) \in \mathcal{V}_\Delta} \sum_{n \in \mathcal{V}} \Delta(u_n^{(k)}, u_n^{(j)}) \\ &= \sum_{q \in \mathcal{V}'} \Psi(u_q) + \sum_{(q,r) \in \mathcal{E}'} \Theta(u_q, u_r), \end{aligned} \tag{3}$$

where $\mathcal{V}_\Delta$ is the edge set for all diversity terms in the MSG, the node set $\mathcal{V}'$ is composed of the nodes from the K basic subgraphs, the edge set $\mathcal{E}'$ includes the edges in the K basic subgraphs as well as the edges for the diversity terms between any two basic subgraphs, and $\Theta(\cdot, \cdot)$ is the combination of all smoothness terms $\Phi(\cdot, \cdot)$ and all diversity terms $\Delta(\cdot, \cdot)$. Note that our MSG energy function in Eq. (3) shares the same formulation as the standard graph. So it can be solved directly by existing energy minimization methods to yield all of the multiple states at once.

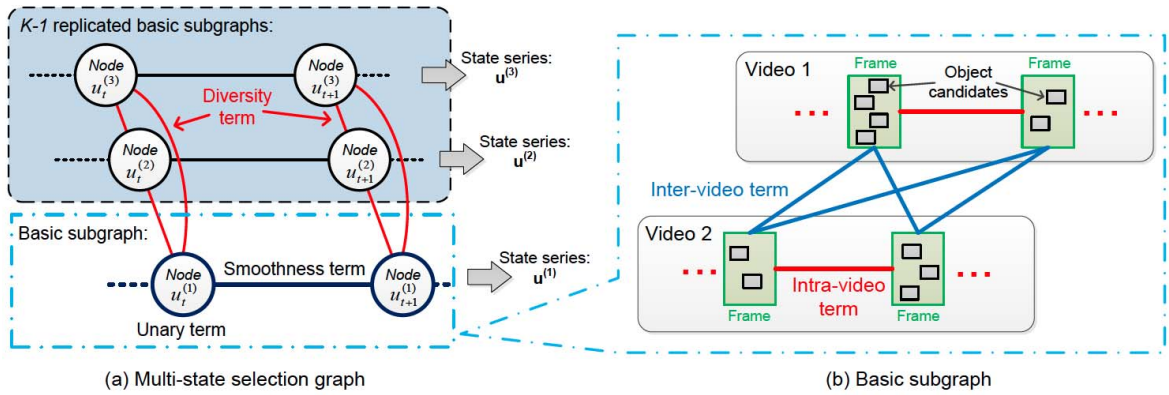


Fig. 2. (a) Our multi-state selection graph (MSG), illustrated for $K = 3$. For K-state selection, our method replicates the basic subgraph $K-1$ times to form K basic subgraphs, and connects the corresponding nodes of different basic subgraphs with the diversity term. Each basic subgraph outputs its corresponding candidate series. (b) Our basic subgraph for video co-segmentation. In this subgraph, each frame of one video is a node, and the foreground object candidates of the frame are the states that this node can take. The nodes (frames) from different videos are fully connected by the inter-video terms. Within a given video, only adjacent nodes (frames) are connected by the intra-video terms. Our basic subgraph can be directly employed in our MSG, where the smoothness term of the MSG includes the inter-video and intra-video terms.

IV. Object-Based Video Co-Segmentation

In this section, we employ our MSG to solve the object-based multiple foreground video co-segmentation problem. Suppose we have V videos, and each video consists of T_v frames. In each frame, a set of (possibly overlapping) object candidates is generated by using [4]. To identify the foreground object in each frame, we examine various region properties that reflect the characteristics of foregrounds, while accounting for intra-video coherence of the foreground as well as foreground consistency among the different videos. Thus we provide a basic subgraph for video co-segmentation, in which each video is modeled as a sequence of nodes, illustrated in Fig. 2(b). In this graph, each node represents a frame in the video, and the possible states of a node are the set of foreground object candidates in the frame. Concatenating the selected candidates from all the frames of videos in the set gives a *candidate series* $\mathbf{u} = \{u_{v,t} | v = 1, ..., V; t = 1, ..., T_v\}$. For each video, *intra-video* edges are placed between the nodes of adjacent frames. In addition, the nodes of different videos are fully connected with each other by *inter-video* edges.

A. MSG for Video Co-Segmentation

Our basic subgraph for video co-segmentation can be directly employed in our MSG graph for the purpose of multiple object co-segmentation as discussed in our preliminary conference paper [9]. However, our MSG selects the same number of states for each node. This means that for multiple object video co-segmentation, the objects must appear in each frame of each video. This is a strong assumption that cannot hold in general. When a foreground object is missing in one of the videos, our MSG will segment an unrelated region in that video. To avoid this problem, we introduce an indicator matrix that handles missing foreground objects. Let $\mathbf{Y} \in \mathbb{R}^{V \times K}$ denote the indicator matrix for video co-segmentation, where each entry y_{vk} indicates whether video v contains object k. If it does, then $y_{vk} = 1$. Otherwise, $y_{vk} = 0$. With the indicator matrix $\mathbf{Y}$, we formulate the energy function of our MSG for video co-segmentation as follows:

$$
\begin{aligned}
E_{cs} = & \sum_{k=1}^{K} \sum_{v=1}^{V} \sum_{t=1}^{T_v} y_{vk} \left[\Psi(u_{v,t}^{(k)}) + \Phi_\alpha(u_{v,t}^{(k)}, u_{v,t+1}^{(k)}) \right] \\
& + \sum_{k=1}^{K} \sum_{\substack{v,w=1, \\ v \neq w}}^{V} \sum_{t=1}^{T_v} \sum_{s=1}^{T_w} y_{vk} y_{wk} \Phi_\beta(u_{v,t}^{(k)}, u_{w,s}^{(k)}) \\
& + \sum_{\substack{k,j=1 \\ k \neq j}}^{K} \sum_{v=1}^{V} \sum_{t=1}^{T_v} y_{vk} y_{vj} \Delta(u_{v,t}^{(k)}, u_{v,t}^{(j)}),
\end{aligned} \tag{4}
$$

where $u_{v,t}^{(k)}$ denotes the k^{th} selected candidate in frame t of video v, and $\Psi(\cdot)$, $\Phi_\alpha(\cdot,\cdot)$, $\Phi_\beta(\cdot,\cdot)$, and $\Delta(\cdot,\cdot)$ represent the unary, intra-video, inter-video and diversity terms, respectively. These terms are defined in the following subsection, and are also discussed in [9]. We note that if all elements in the indicator matrix $\mathbf{Y}$ are set to 1, the energy function in Eq. (4) reduces to the one in [9]. The energy function can be further reduced to that of a graph for single object co-segmentation, discussed in [9], by setting all elements in the indicator matrix $\mathbf{Y}$ to 1 and $K = 1$.

B. Term Definitions

Unary term $\Psi(\cdot)$ combines three factors for determining the likelihood that an object candidate is the foreground:

$$
\Psi(u) = -\log\left[O(u) \cdot \max\left(M(u), S(u)\right)\right]. \tag{5}
$$

The factors that influence this energy are the objectness score $O(u)$, motion score $M(u)$, and saliency score $S(u)$ of the candidate u. The objectness score $O(u)$ is taken to be the value returned by the candidate generation process [4]. The motion score $M(u)$ represents the confidence that candidate u corresponds to a coherently moving object in the video, and has the same definition as in [6] and [9]. In our paper, we also consider the static saliency cue, since the foreground object may not always be moving in the video.

Co-saliency detection is different from saliency detection for single images, in that it discovers common saliency among multiple images [37]–[40]. A video can be treated as a sequence of images, with the foreground highlighted in a co-saliency map. Moreover, the common foreground in multiple videos can also be determined based on co-saliency. In our work, we compute the co-saliency map for each frame by using [38], and then calculate the mean co-saliency value for each candidate region as the saliency score $S(u)$.

Intra-video term $\Phi_\alpha(\cdot,\cdot)$ provides a spatio-temporal smoothness constraint between neighboring frames in an individual video. We define this term as follows:

$$\Phi_\alpha(u,u') = \gamma_1 \cdot D_c(u,u') \cdot D_f(u,u'), \tag{6}$$

where u and u' are any two candidates from neighboring frames that are not necessarily consecutive, γ_1 is a weighting coefficient, $D_c(\cdot,\cdot)$ represents the χ^2-distance between color histograms, and $D_f(\cdot,\cdot)$ represents the overlap between the two candidates in adjacent frames:

$$D_f(u,u') = -\log\left(\frac{|u \cap Warp(u')|}{|u \cup Warp(u')|}\right), \tag{7}$$

where $Warp(u')$ transforms the candidate region u' from its frame to the frame of u based on optical flow mapping [41].

Inter-video term $\Phi_\beta(\cdot,\cdot)$ measures foreground consistency among the videos. In the basic subgraph, candidates from one video are connected to those in the other videos. The inter-video energy is defined as

$$\Phi_\beta(u,u') = \gamma_2 \cdot D_c(u,u') \cdot D_s(u,u'), \tag{8}$$

where u and u' are any two candidates from different videos, γ_2 is a weight, and $D_c(\cdot,\cdot)$ and $D_s(\cdot,\cdot)$ respectively denote the χ^2-distances of color histograms and HOG descriptors [42] for the region within a minimum bounding box enclosing the candidate.

Diversity term $\Delta(\cdot,\cdot)$ is used to avoid selecting the same object in different candidate series. Here, we define the diversity term as the intersection-over-union metric between two candidates:

$$\Delta(u^{(k)},u^{(j)}) = \gamma_3 \frac{|u^{(k)} \cap u^{(j)}|}{|u^{(k)} \cup u^{(j)}|}, \tag{9}$$

where γ_3 is a scale parameter.

C. Optimization Procedure

As mentioned above, if all elements in the indicator matrix $\mathbf{Y}$ are set to 1, the energy function in Eq. (4) reduces to the one in [9]. Alternatively, in another special case called the weakly supervised case, the objects that exist in each video are known, and the indicator matrix $\mathbf{Y}$ can thus be set accordingly. In these two cases, the indicator matrix $\mathbf{Y}$ is fixed. So the energy function in Eq. (4) shares the same formulation as that in Eq. (3), which can be solved directly by using existing energy minimization methods to yield all of the multiple states.

Let us consider the more general case, where the indicator matrix $\mathbf{Y}$ is unknown. In this case, the energy function of our MSG in Eq. (4) contains two sets of variables: the indicator matrix $\mathbf{Y}$ and the selected candidate series $\{\mathbf{u}^{(1)},\ldots,\mathbf{u}^{(K)}\}$. To solve the non-trivial optimization problem in Eq. (4), we develop an iterative procedure to approximately update these two sets of variables until convergence. Our multiple foreground video co-segmentation method is explained below.

1) Initializing the Indicator Matrix: All the values in the indicator matrix $\mathbf{Y}$ are initialized to 1, and then updated in subsequent iterations.

2) Solving the Candidate Series $\{\mathbf{u}^{(1)},\ldots,\mathbf{u}^{(K)}\}$*:* With the fixed indicator matrix $\mathbf{Y}$, the energy function in Eq. (4) can be represented as the standard graph in Eq. (3), which can be solved directly by using existing energy minimization methods. In this paper, TRW-S [36] is employed to solve the MSG energy.

3) Updating the Indicator Matrix: With the candidate series $\{\mathbf{u}^{(1)},\ldots,\mathbf{u}^{(K)}\}$ obtained in the previous step, we can update the indicator matrix. As mentioned before, the method discussed in our preliminary conference paper [9] will segment an irrelevant region in a video if the foreground object is missing. Generally, these irrelevant regions have relatively lower unary scores than the actual foregrounds. Thus we can update the indicator matrix $\mathbf{Y}$ based on the unary scores of $\{\mathbf{u}^{(1)},\ldots,\mathbf{u}^{(K)}\}$ from the previous iteration. Let us denote y_{vk} as the indicator value of foreground k in video v. We update y_{vk} as follows:

$$y_{vk} = \begin{cases} 1, & \text{if } \bar{\Psi}(\mathbf{u}_v^{(k)}) \leqslant \tau^k \\ \varepsilon, & \text{otherwise,} \end{cases} \tag{10}$$

where $\mathbf{u}_v^{(k)}$ denotes the selected candidate series $\{u_{v,t}^{(k)}\}_{t=1}^{T_v}$ of foreground k in video v, and $\bar{\Psi}(\mathbf{u}_v^{(k)})$ is the average unary score of the selected candidate series $\mathbf{u}_v^{(k)}$, defined as

$$\bar{\Psi}(\mathbf{u}_v^{(k)}) = \frac{1}{T_v}\sum_{t=1}^{T_v}\Psi(u_{v,t}^{(k)}). \tag{11}$$

τ^k denotes the global unary score for foreground k in all videos, and is defined as

$$\tau^k = \frac{1}{V}\sum_{v=1}^{V}\bar{\Psi}(\mathbf{u}_v^{(k)}). \tag{12}$$

With this threshold, irrelevant segments are distinguished and the indicator matrix is updated according to Eq. (10). Note that if the indicator y_{vk} is updated to 0, it would henceforth generate irrelevant regions and have no chance of being updated back to 1. To avoid this problem in the unsupervised method, we set y_{vk} to ε instead of 0, where ε is a small value ($\varepsilon = 0.001$ in our implementation). As a result, the basic subgraph for state series $\mathbf{u}_v^{(k)}$ still has minimal influence on the global energy function in Eq. (4), but our algorithm will select a corresponding region if one actually exists. This would allow y_{vk} to be updated back to 1 in Eq. (10) if foreground k was mistakenly considered as missing in video v in an earlier iteration.

4) Stopping Criterion: The iterations can be stopped when the indicator matrix does not change or upon reaching a predefined maximum number of iterations. In our experiments, our optimization algorithm terminates in 3-4 iterations for our video co-segmentation task.

5) Pixel-Level Post-Process: The segmentation of object candidates in [4] is imprecise, due to the superpixel extraction and merging phases in proposal generation. For each video, we thus follow [1], [2] to refine the selected objects through a pixel-level post-process by using a spatiotemporal graph-based segmentation method.

V. Experiments

In our preliminary conference paper [9], we demonstrated that our MSG outperforms several existing approaches [2], [3], [8], [21] for single foreground co-segmentation. Here, we conduct the experiment using the MOViCS dataset [3] for multiple foreground video co-segmentation. MOViCS contains four video sets with 11 videos in total. Five frames of each video are labeled with ground truth at the pixel level. We also construct a new multiple foreground video dataset, composed of four sets of video pairs each with two foreground objects in common. The ground truth is manually labeled for each frame.

Evaluation Measures: We employ two metrics for the multi-object co-segmentation task. The first is the *average per-frame pixel error* defined as

$$Score_1 = \frac{1}{K}\sum_{k=1}^{K}\left[\min_i \frac{|\mathrm{XOR}(R_i, GT_k)|}{T_{all}}\right], \tag{13}$$

where K is the number of foreground objects, R_i denotes all segmented foreground results for all frames in the video set that are from foreground class i, and GT is the ground truth. T_{all} is the total number of frames in the video set. The second measure is the *average Intersection-over-Union (IoU) metric*, which is defined as

$$Score_2 = \frac{1}{K}\sum_{k=1}^{K}\left[\max_i \frac{|R_i \cap GT_k|}{|R_i \cup GT_k|}\right]. \tag{14}$$

A. Experiments on the MOViCS Dataset

We first test our method on the MOViCS dataset [3]. We compare our method with three multi-object co-segmentation methods:

- **Multi-class image co-segmentation (MIC)** [21], which segments multiple images into regions of multiple classes. The class that most overlaps the ground truth over the video set is selected as the foreground segmentation result. Since MIC requires a predefined number of clusters K, we sample values of K between 5 and 8, and choose the value that yields the best performance for each video set.
- **Multi-class video co-segmentation (MVC)** [3], which produces a segmentation of multiple classes from multiple videos. As done for MIC [21], the segmentation result is taken as the class with the most overlap to the ground truth over the entire video set.
- **Regulated Maximum Weight Cliques (RMWC)** [14], which employs the object proposal tracklet as the basic element, and uses the regulated maximum weight clique method to select the corresponding nodes for video multi-class segmentation.

TABLE I

Average Per-Frame Pixel Errors on the MOViCS Dataset. The Primary Object Is Used to Denote the Corresponding Set

Methods	Chicken	Giraffe	Lion	Tiger	Avg.
MIC [21]	8957	8291	13654	44809	18928
MVC [3]	5158	2476	6419	34352	12101
RMWC [14]	2448	2958	6459	47316	14720
Our MSG_s	4343	4613	9679	21005	9910
Our MSG_u	2372	3396	11828	21005	9650
Our MSG_w	2372	3396	6084	21005	8214

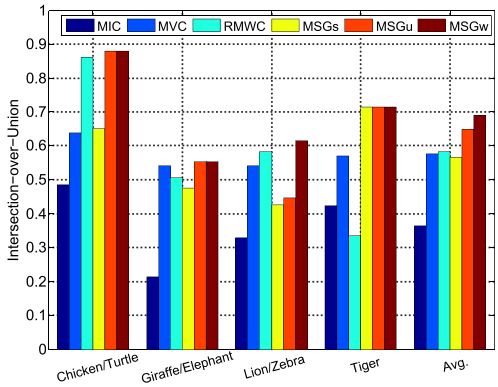


Fig. 3. IoU metric on the MOViCS dataset.

For our method, three different configurations are examined. The first is the MSG method discussed in our preliminary conference version [9] (referred to as MSG_s), in which all elements in the indicator matrix $\mathbf{Y}$ are set to 1. The second is weakly supervised co-segmentation (referred to as MSG_w), in which the indicator matrix $\mathbf{Y}$ is set directly based on which objects exist in each video, as described in Sec. IV-C. The last one is our unsupervised foreground co-segmentation method (referred to as MSG_u), where the indicator matrix $\mathbf{Y}$ is learned by using the iterative procedure. For all methods, the default object number is set to $K = 2$, except for the Tiger set which only contains one object (i.e., we set $K = 1$ in this set). Fig. 4 displays multiple foreground segmentation results on the MOViCS dataset, and quantitative results are given in Table I and Fig. 3.

Local appearance and spatial consistency terms are combined with class-level discrimination in MIC [21]. Since MIC is designed for image co-segmentation, it does not include a temporal smoothness constraint for the video. Also, a low-level representation without an objectness constraint is employed to classify pixels, which may lead to incorrectly merged object classes from the foreground and background. For example, the elephant in the first video of the Giraffe/Elephant set is wrongly classified together with the background bushes in the second video. For complex frames such as the second video of the Lion/Zebra set, MIC produces a segment (outlined in yellow) that contains multiple objects (lion and zebra) and the background (bushes).

MVC [3] includes a temporal smoothness constraint and obtains better performance than MIC. However, the pixel-level processing of MVC leads to class labeling errors and hence to incorrect correspondences of objects (e.g., different labels of the zebras in the second and fourth videos of the Lion/Zebra set).

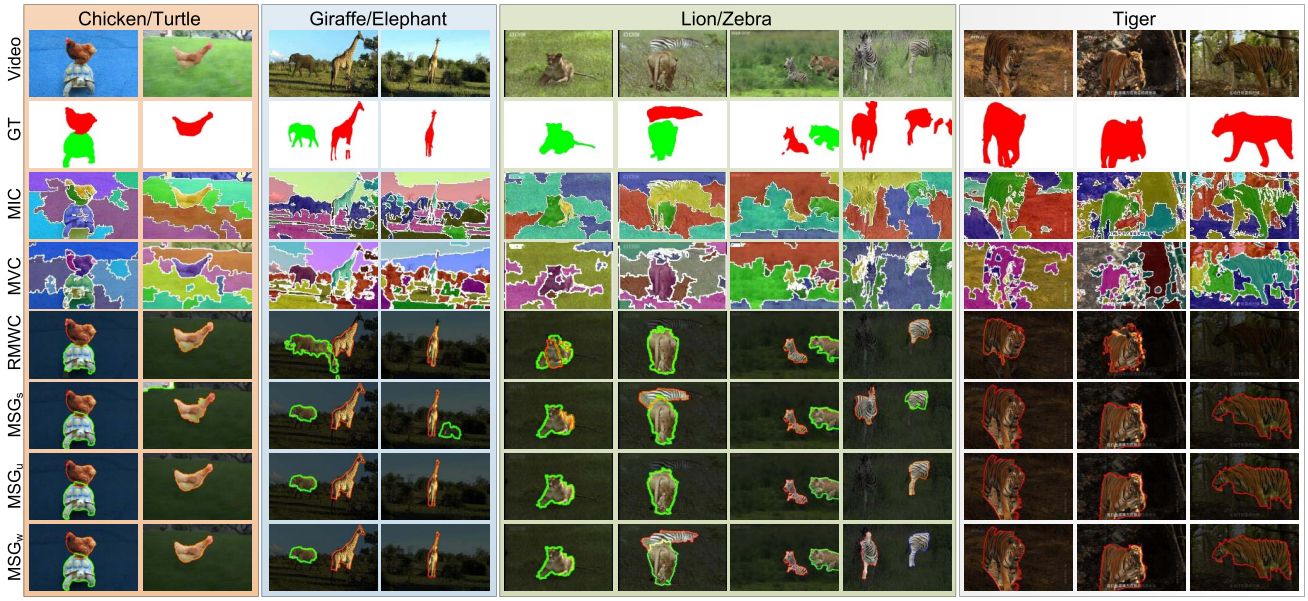


Fig. 4. Multiple foreground co-segmentation on the MOViCS dataset. From top to bottom: input videos, ground truth, MIC [21], MVC [3], RMWC [14], and our methods MSG$_s$, MSG$_u$ and MSG$_w$.

The object-based methods (RMWC [14] and our MSG) avoid the aforementioned issues and provide more meaningful foreground co-segmentation results. RMWC [14] utilizes a tracklet similarity threshold to connect corresponding tracklets. However, the fixed threshold is not robust to appearance variations, and may lead to missing objects. For example, the illumination difference in the last video of Tiger causes a different object appearance compared to the other two videos. As a result, RMWC [14] cannot find the corresponding object in the last video.

Our MSG$_s$ discussed in the preliminary conference paper [9] produces the same number of regions in each video, which may result in irrelevant regions being segmented in the videos that are missing one or more of the foreground objects. For example, the second video of the Giraffe/Elephant set in Fig. 4 has no elephant. MSG$_s$ nevertheless segments a secondary region (outlined in green) that is most similar to the elephant in the first video, but is not an elephant. Moreover, in the last video of the Lion/Zebra set, the two zebras are treated independently of each other since our method assumes that each video contains up to one instance of each foreground class. As a result, the zebra on the right in the last video is not considered as part of the zebra, but rather is co-segmented as part of the lion class, which lacks an actual instance in this video.

By contrast, our methods MSG$_u$ and MSG$_w$ provide stable segmentation results, with the weakly supervised MSG$_w$ performing the best. MSG$_u$ obtains the same co-segmentation results as MSG$_w$ on most of the MOViCS examples because the indicator matrix is estimated correctly. An exception is the Lion/Zebra set in Fig. 4, where our MSG$_u$ segments the lion successfully, but mis-segments the zebra in the second and fourth videos. A possible reason is that the zebra is naturally camouflaged among the background bushes, which makes its unary score higher than the global threshold for the zebra in Eq. (10). Although this mis-segmentation of our unsupervised MSG$_u$ method leads to more errors than MVC [3] for this Lion/Zebra set, our MSG$_u$ can generally segment a whole object as an independent region. As shown in Fig. 4, MVC [3] produces many fragments in the segmented region, such as the lion in the first video and the zebra in the third video. By contrast, our MSG$_u$ generally preserves the entire object in each segmented region. Moreover, our MSG$_u$ performs better than MIC [21], MVC [3] and RMWC [14] on average over the whole MOViCS dataset.

As mentioned above, our method assumes that each video has up to one instance for each foreground class. When there are multiple instances of the same class in one video, e.g. the last video in the Lion/Zebra set, the MSG$_u$ method outputs only one zebra. In the weakly supervised approach MSG$_w$, this problem can be easily solved by using an additional foreground class, such as by setting the foreground number to $K = 3$ in the Lion/Zebra set, with one for lion and two for zebras. The secondary zebra in the last video is then segmented as a separate object.

B. Experiments on Our Multiple Foreground Video Dataset

The second dataset is our multiple foreground video dataset. With this dataset, our MSG$_u$ obtains the same indicator matrix as the weakly supervised method MSG$_w$, and achieves the same results as MSG$_s$ and MSG$_w$. Thus we report only a single set of results for them, denoted as MSG in this experiment. We also report the results of three multi-class co-segmentation methods, MIC [21], MVC [3] and RMWC [14]. Fig. 5 displays multiple foreground segmentation results on our dataset, and quantitative results are reported in Table II and Fig. 6.

Like with the MOViCS dataset, our method outperforms the other methods on the most videos of our multiple foreground video dataset. As the indicator matrix is only defined on each independent video, within a video our MSG method

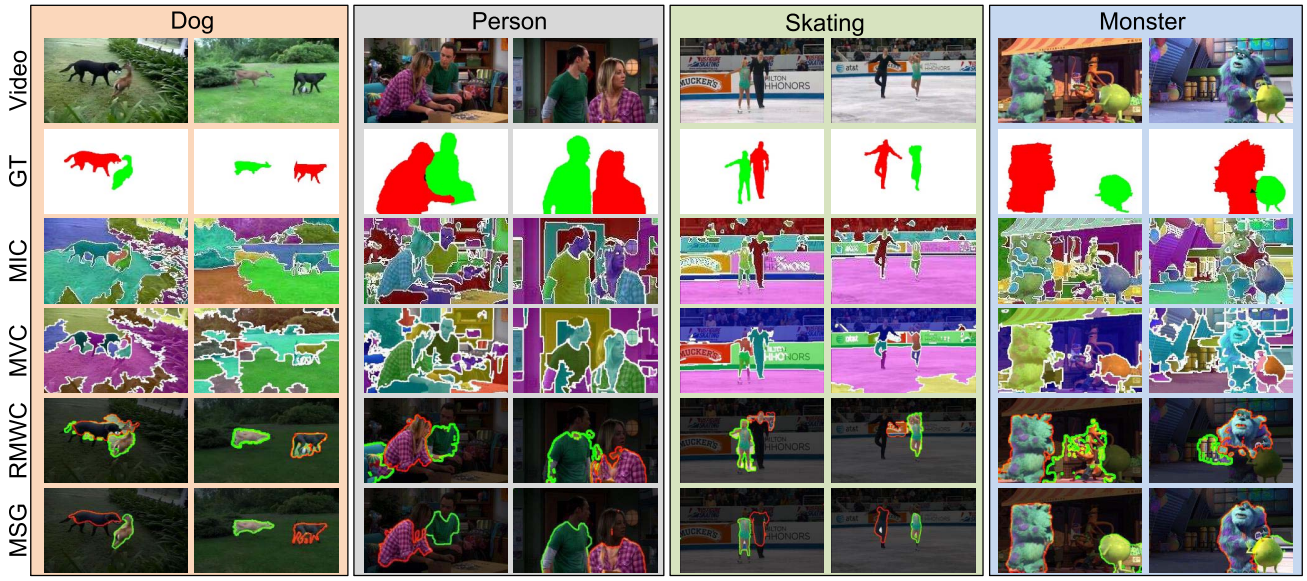


Fig. 5. Segmentation results on our newly collected multiple foreground video dataset. From top to bottom: input videos, ground truth, MIC [21], MVC [3], RMWC [14], and our MSG method.

TABLE II
AVERAGE PER-FRAME PIXEL ERRORS ON OUR MULTIPLE FOREGROUND VIDEO DATASET

Methods	Dog	Person	Monster	Skating	Avg.
MVC [3]	1807	10389	7394	10223	7453
MIC [21]	4794	11033	7836	26616	12570
RMWC [14]	8071	10736	10109	31770	15171
Our MSG	1115	9321	3551	3274	4315

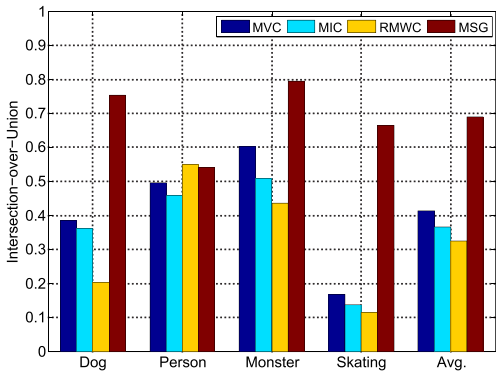


Fig. 6. IoU metric on our multiple foreground dataset.

still follows the common constraint as in single video segmentation [1], [2], that the foreground of a video should appear in each frame. This constraint is beneficial for foreground extraction, and helps our method to achieve better results. Even when the foreground is not visible in a few frames, we can still use the pixel-level segmentation in the post-processing step to solve this problem.

C. Running Time

The main running time costs of our method are from object candidate generation and feature extraction (about 4 mins/image), which can easily be accelerated through parallel computing. Besides this, the major cost arises from our MSG, including graph building and solving by TRW-S [36]. Based on the current sizes of publicly available video datasets used for video co-segmentation, we additionally perform two experiments on our workstation with an Intel Xeon 3.2GHz CPU and 16GB RAM. In the first experiment, we process two 2-minute videos (about 700 frames each) with two objects, which takes our MSG less than 320s. The second experiment has 40 short videos (averaging 50 frames each) with two objects, and it takes about 442s. These experiments demonstrate the scalability of our method on long/large video collections.

D. Discussion

1) Relationship With Object Proposals: Our object-based segmentation is based on object proposal generation. Here, we discuss the relationship between object proposals and our method. First, object proposal generation methods (see [4], [16]) produce a pool of segment proposals that have been used for object detection. These proposal generation methods can detect most image object regions with high accuracy. We also observe that the generated proposals can reasonably cover most objects in our experiments.

Second, object-based segmentation methods ([2], [5], [6], [14]) and our method are all based on object proposals, and thus inherit the limitations of these object proposals. For example, when an object is composed of multiple components, the object proposals may be composed of multiple different sub-regions, such as for the person's face in Fig. 5. We must however note that the pixel-based methods ([3], [21]) also fail in this case, as they would typically segment the components into different clusters.

Third, our method formulates a multi-state selection graph (MSG) and incorporates the inter-video and intra-video constraints in the selection of the correct foreground objects

from cluttered backgrounds. In our experiments, we observe that most correctly selected objects are not at the top of the proposal pool. For multiple object cases, the secondary object is often ranked about 5-10.

2) Number of Objects: In the most existing image and video co-segmentation methods [10], [11], [14], [20], [21], the number of objects K, which determines how many objects are to be segmented, must be provided beforehand based on prior information. When the given K is smaller than the number of common foreground objects in all videos in the set, our algorithm can still select the top K objects from all object candidates.

VI. Conclusion

In this paper, we proposed an object-based video co-segmentation method that can extract multiple foreground objects in a video set. The key contributions of this work are the multi-state selection graph (MSG) to address the problem of multiple foreground objects, and introduction of the indicator matrix. The latter removes the constraint that every foreground object (to be co-segmented) must be present in all the videos and thus prevents the mis-segmentation of irrelevant regions as the missing objects. Our MSG provides a general and global framework that can be used for extending any standard graph model to handle multiple state selection while still allowing optimization by existing energy minimization techniques. With the inclusion of the indicator matrix, MSG becomes more flexible and practical for video co-segmentation of multiple foregrounds, since any common foreground object in a video set does not necessarily have to appear in every video.

References

[1] T. Ma and L. J. Latecki, "Maximum weight cliques with mutex constraints for video object segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 670–677.

[2] D. Zhang, O. Javed, and M. Shah, "Video object segmentation through spatially accurate and temporally dense extraction of primary object regions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 628–635.

[3] W.-C. Chiu and M. Fritz, "Multi-class video co-segmentation with a generative multi-video model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 321–328.

[4] I. Endres and D. Hoiem, "Category-independent object proposals with diverse ranking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 222–234, Feb. 2014.

[5] S. Vicente, C. Rother, and V. Kolmogorov, "Object cosegmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 2217–2224.

[6] Y. J. Lee, J. Kim, and K. Grauman, "Key-segments for video object segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1995–2002.

[7] H. Fu, D. Xu, S. Lin, and J. Liu, "Object-based RGBD image co-segmentation with mutex constraint," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 4428–4436.

[8] F. Meng, H. Li, G. Liu, and K. N. Ngan, "Object co-segmentation based on shortest path algorithm and saliency model," *IEEE Trans. Multimedia*, vol. 14, no. 5, pp. 1429–1441, Oct. 2012.

[9] H. Fu, D. Xu, B. Zhang, and S. Lin, "Object-based multiple foreground video co-segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3166–3173.

[10] D.-J. Chen, H.-T. Chen, and L.-W. Chang, "Video object cosegmentation," in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 805–808.

[11] J. C. Rubio, J. Serrat, and A. López, "Video co-segmentation," in *Proc. Asian Conf. Comput. Vis.*, 2012, pp. 13–24.

[12] J. Guo, Z. Li, L.-F. Cheong, and S. Z. Zhou, "Video co-segmentation for meaningful action extraction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2232–2239.

[13] J. Guo, L.-F. Cheong, R. T. Tan, and S. Z. Zhou, "Consistent foreground co-segmentation," in *Proc. Asian Conf. Comput. Vis.*, 2014, pp. 241–257.

[14] D. Zhang, O. Javed, and M. Shah, "Video object co-segmentation by regulated maximum weight cliques," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 551–566.

[15] X. Cao, F. Wang, B. Zhang, H. Fu, and C. Li, "Unsupervised pixel-level video foreground object segmentation via shortest path algorithm," *Neurocomputing*, 2015, to be published.

[16] J. Carreira and C. Sminchisescu, "CPMC: Automatic object segmentation using constrained parametric min-cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1312–1328, Jul. 2012.

[17] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, 2013.

[18] A. Ion, J. Carreira, and C. Sminchisescu, "Probabilistic joint image segmentation and labeling by figure-ground composition," *Int. J. Comput. Vis.*, vol. 107, no. 1, pp. 40–57, 2013.

[19] G. Kim, E. P. Xing, L. Fei-Fei, and T. Kanade, "Distributed cosegmentation via submodular optimization on anisotropic diffusion," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 169–176.

[20] G. Kim and E. P. Xing, "On multiple foreground cosegmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 837–844.

[21] A. Joulin, F. Bach, and J. Ponce, "Multi-class cosegmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 542–549.

[22] L. Mukherjee, V. Singh, J. Xu, and M. D. Collins, "Analyzing the subspace structure of related images: Concurrent segmentation of image sets," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 128–142.

[23] R. Jin and Z. Ghahramani, "Learning with multiple labels," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 897–904.

[24] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1819–1837, Aug. 2014.

[25] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 681–687.

[26] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognit.*, vol. 40, no. 7, pp. 2038–2048, 2007.

[27] C. Wang, S. Yan, L. Zhang, and H.-J. Zhang, "Multi-label sparse coding for automatic image annotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1643–1650.

[28] C.-T. Nguyen, D.-C. Zhan, and Z.-H. Zhou, "Multi-modal image annotation with multi-instance multi-label LDA," in *Proc. Int. Joint Conf. Artif. Intell.*, 2013, pp. 1558–1564.

[29] X. Xue, W. Zhang, J. Zhang, B. Wu, J. Fan, and Y. Lu, "Correlative multi-label multi-instance image annotation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 651–658.

[30] J. He, L. Balzano, and A. Szlam, "Incremental gradient on the Grassmannian for online foreground and background separation in subsampled video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1568–1575.

[31] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 597–610, Mar. 2013.

[32] R. Szeliski *et al.*, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1068–1080, Jun. 2008.

[33] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[34] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

[35] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.

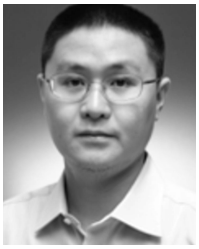
[36] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1568–1583, Oct. 2006.

[37] H. Li and K. N. Ngan, "A co-saliency model of image pairs," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3365–3375, Dec. 2011.

[38] H. Fu, X. Cao, and Z. Tu, "Cluster-based co-saliency detection," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3766–3778, Oct. 2013.

[39] X. Cao, Z. Tao, B. Zhang, H. Fu, and W. Feng, "Self-adaptively weighted co-saliency detection via rank constraint," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4175–4186, Sep. 2014.
[40] X. Cao, Y. Cheng, Z. Tao, and H. Fu, "Co-saliency detection via base reconstruction," in *Proc. ACM Int. Conf. Multimedia*, 2014, pp. 997–1000.
[41] C. Liu, "Beyond pixels: Exploring new representations and applications for motion analysis," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., MIT, Cambridge, MA, USA, 2009.
[42] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 2005, pp. 886–893.

Huazhu Fu received the B.S. degree in mathematical sciences from Nankai University, in 2006, the M.E. degree in mechatronics engineering from the Tianjin University of Technology, in 2010, and the Ph.D. degree in computer science from Tianjin University, China, in 2013. He is currently a Research Fellow with the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include computer vision, image processing, and medical image analysis.



Dong Xu (M'07–SM'13) received the B.E. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 2001 and 2005, respectively.

He was with Microsoft Research Asia, Beijing, China, and the Chinese University of Hong Kong, Hong Kong, for over two years, while pursuing the Ph.D. degree. He was a Post-Doctoral Research Scientist with Columbia University, New York, NY, USA, for one year. He also worked as a Faculty Member with the School of Computer Engineering, Nanyang Technological University, Singapore. He is currently a Faculty Member with the School of Electrical and Information Engineering, The University of Sydney, Australia. His current research interests include computer vision, statistical learning, and multimedia content analysis.

Dr. Xu co-authored a paper that received the Best Student Paper Award in the IEEE International Conference on Computer Vision and Pattern Recognition in 2010. His another co-authored paper also won the IEEE TRANSACTIONS ON MULTIMEDIA Prize Paper Award in 2014.

Bao Zhang received the B.E. degree in software engineering from the School of Computer Software, Tianjin University, Tianjin, China, in 2010, and the M.E. degree from the School of Computer Science and Technology, Tianjin University, in 2013. His current research interests include computer vision, scene classification, video processing, image saliency detection, and segmentation.

Stephen Lin (M'10) received the B.S.E. degree in electrical engineering from Princeton University, Princeton, NJ, USA, and the Ph.D. degree in computer science and engineering from the University of Michigan, Ann Arbor. He is a Senior Researcher with the Internet Graphics Group, Microsoft Research. He served as a Program Co-Chair for the International Conference on Computer Vision in 2011 and is on the Editorial Board of the *International Journal of Computer Vision*. His research interests include computer vision, image processing, and computer graphics.

Rabab Kreidieh Ward (F'99) has around 40 years of post-doctoral experience in academic education, research, development, and leadership. She has authored around 500 papers in refereed journals and conferences and holds six patents. Her research interests are mainly in broad areas of signal and image processing and their applications. Some of her work has been licensed to the U.S. and Canadian industry. She is a fellow of the Royal Society of Canada, the Canadian Academy of Engineers, and the Engineering Institute of Canada. Amongst her large number of awards are the UBC Senior Killam Mentoring Award and UBC Killam Research Prize. She has served and provided leadership to IEEE and other professional societies and is presently the President-Elect of the IEEE Signal Processing Society.